External Validation and Modification of a Predictive Model for Acute Postsurgical Pain at Home after Day Surgery

The Clinical Journal of Pain Publish Ahead of Print
DOI:10.1097/AJP.0000000000000413

OPEN

External validation and modification of a predictive model for acute postsurgical pain at home after day surgery

Björn Stessel, MD[1,2]; Audrey A.A. Fiddelers, PhD[2,3]; Marco A. Marcus MD, PhD[2,4]; Sander M.J. van Kuijk, PhD[5]; Elbert A.Joosten, PhD [2]; Madelon L. Peters, PhD[6]; Wolfgang F.F.A. Buhre, MD, PhD[2]; Hans-Fritz Gramke, MD, PhD[2]

[1]Department of Anesthesiology and Pain Medicine, Jessa Hospital, Hasselt, Belgium

[2]Department of Anesthesiology and Pain Medicine, Maastricht University Medical Center+, Maastricht, the Netherlands

[3]Network Acute Care Limburg, Maastricht, The Netherlands

[4]Department of Anesthesia/ICU, Pain & Palliative Care, Hamad Medical Corporation, Doha, Qatar

[5]Department of Clinical Epidemiology and Medical Technology Assessment, Maastricht University Medical Center+, Maastricht, the Netherlands

[6]Department of Clinical Psychological Science, Maastricht University Medical Center+, Maastricht, the Netherlands.

The authors declare no conflict of interest.

Corresponding author:

B. Stessel

Dep. of Anesthesiology and Pain Treatment

Jessa Hospital – Hasselt, Virga-Jesse Campus

Stadsomvaart 11

3500 Hasselt, Belgium

Phone +32 11289111

Fax +32 11274590

bjornstessel@hotmail.com

The authors declare no conflict of interest.

Abstract

Objectives –

In 2009, Gramke et al. have described predictive factors to pre-operatively detect those at risk for moderate to severe acute postoperative pain (APSP) after day surgery.

The aim of the present study is to externally validate this initial model and to improve and internally validate a modified version of this model.

Methods –

Elective patients scheduled for day surgery were prospectively enrolled from November 2008 to April 2010. Model discrimination was quantified using the area under the receiver operating characteristic curve (AUC). Model calibration was assessed by visual inspection of the calibration plot.

Subsequently, we modified (different assignment of type of surgery, different cut-off for moderate to severe APSP, continuous of dichotomized variables and testing of additional variables) and internally validated this model by standard bootstrapping techniques.

Results –

A total of 1118 patients were included. The AUC for the original model was 0.81 in the derivation dataset and 0.72 in our validation dataset. The model showed poorly calibrated risk predictions . The AUC of the modified model was 0.82 (optimism-corrected AUC = 0.78). This modified model showed good calibration.

Conclusion -

The original prediction model of Gramke et al. performed insufficient on our cohort of outpatients with respect to discrimination and calibration. Internal validation of a modified model shows promising results. In this model, preoperative pain, patient derived expected pain and different types of surgery are the strongest predictors of moderate to severe APSP after day surgery.

Funding

The study was funded solely by the Dept. of Anesthesiology and Pain Medicine, Jessa Hospital, Hasselt, Belgium and the Dept. of Anesthesiology and Pain Medicine, Maastricht University Medical Centre+, Maastricht, the Netherlands.

Keywords

Day surgery

Postoperative pain

Predictive model

Validation

Modification

List of abbreviations

APSP = Acute postsurgical pain

CPSP = Chronic postsurgical pain

CP = Cut-off Point

QOL = Quality of Life

PCS = Pain Catastrophizing Scale

LOT-R = Life Orientation Test Revised

SFQ = Surgical Fear Questionnaire

AUC = Area Under the Curve

ASA PS = American Society of Anesthesiologists Physical Status

Introduction

Despite increased awareness and improvements in postoperative pain management over the last decades, the prevalence of outpatients suffering moderate to severe acute postsurgical pain (APSP) at home still remains high and varies from 9-40% [1-5]. Particularly in the ambulatory setting, good postoperative analgesia is challenging because patients are responsible for controlling their pain at

home by themselves [6] and the types of analgesics (i.e. no strong opioids) as well as the route of administration (i.e. no epidural, intravenous, subcutaneous or intramuscular route) is limited compared to the inpatient setting.

Obviously, identification of patients at increased risk for APSP provides new opportunities: Tailored pain therapy to specific patient needs, assistance with coping and planned overnight stay can prevent the development of prolonged moderate to severe pain. Therefore, Gramke et al. have identified predictive factors for the development of moderate to severe APSP after day surgery [6]. This model however, was not validated in a new dataset.

Before considering use and implementation of a prediction model, the generalizability of the model needs to be evaluated in a new population by external validation [7]. External validation may be performed by either (partly) the same authors or by completely different teams [8]. Furthermore, the dataset can be retrieved either in the same center (i.e. temporal validation) or in a different one (independent validation)[7].To assess the performance of a previously described prediction model on a new dataset, predicted and observed risks should be compared (i.e. calibration) and the ability of the model to differentiate between patients with or without the event of interest should be quantified (i.e. discrimination) [7,9,10]. Unfortunately, external validation of predictive models is still very uncommon [8], but highly desirable.

Hence, the primary objective of this study is to externally validate a previously described predictive model of APSP after ambulatory surgery [6].

In this model [7], different types of surgery were assigned into two groups according to anticipated level of postoperative pain (i.e. minor or intermediate). In recent years it has been advocated to assign types of surgery to a wide range of surgical procedures (or groups of closely related procedures) as subdivision into relatively broad categories is not precise [11]. Furthermore, in the previously described model [7] the cut-off point (CP) for moderate to severe pain was set on a Visual Analogue Scale (VAS 0-100) > 40 and the variables were dichotomized. However, recent studies have identified a threshold of Numeric Rating Scale (NRS 0-10) > 3 between mild and moderate-to-severe postoperative pain [12,13].

Therefore, the second aim of this study is to modify the previously described prediction model of APSP after ambulatory surgery [7], not only by assigning the types of surgery to a wide range of surgical procedures (or groups of closely related procedures), but also by setting the CP for moderate to severe pain on an NRS > 3. Furthermore, continuous variables instead of dichotomized variables were used and the predictive power of additional variables, like American Society of Anesthesiologists physical status classification (ASA-level), work status, preoperative analgesic use and baseline quality of life (QOL) was included in the analysis.

Materials and methods

Patients

A prospective longitudinal cohort study was used for external validation and modification of a previously published prediction model [6]. The study was approved by the institutional Ethics Committee of the Maastricht University Medical Center+ in 2008, and all patients gave informed consent to participate. All patients undergoing day surgery were eligible to participate, regardless of the type of surgery. Exclusion criteria were 1. patients age <18 years, 2. inability to express themselves, 3. visual dysfunction, or 4. insufficient understanding of the Dutch language.

Questionnaires

Patients were asked to complete two successive questionnaire packages.

First, a baseline questionnaire package was used to measure demographics (e.g. age, gender, educational level, work status, highest level of education), average and present pain intensity, expected postoperative pain intensity by the patient, preoperative analgesic use, previous surgery (related or not) and baseline quality of life (QOL). The EuroQol (EQ-5D) questionnaire was used to analyze QOL [14]. All questions regarding pain were measured on an 11-point Numeric Rating Scale (NRS; where 0 = no pain, and 10 = worst pain imaginable). Furthermore, psychological variables (i.e. catastrophic thinking, personality trait optimism, fear of potential short and long term consequences of surgery) were analyzed based on three validated questionnaires: the Pain Catastrophizing Scale (PCS), the Life Orientation Test Revised (LOT-R) and Surgical Fear Questionnaire (SFQ) [15-18]. For the PCS and LOT-R, shortened versions were used to diminish patient burden [15,19]. In the PCS questionnaire, six of the

thirteen original items were used. These were two questions loading highest on each of the three subscales (i.e. Items 5 and 12 for Helplessness, Items 9 and 11 for Rumination and Items 6 and 13 for Magnification)[15,16]. In the LOT-R, four of the originally ten questions were used. Four filler questions were omitted and the four questions (two positively phrased and two negatively phrased) loading highest on respectively the optimism and pessimism factor were selected [15,19].

Second, a follow-up questionnaire was used to measure APSP related to the surgery on an 11-point NRS.

Procedure

Between November 2008 and April 2010, patients planned for day surgery and presenting at the outpatient clinic for preoperative assessment at the Maastricht University Medical Center+, were asked to participate. If consent was obtained, the patient received an envelope containing an informative letter about the study, the two questionnaire packages and two return envelopes. Patients were instructed to complete the baseline questionnaire package one week before the surgical procedure. Patients who did not return this questionnaire package were considered to be unwilling to participate and no further attempts to contact them were made. The follow-up questionnaire package had to be completed at the fourth day after the surgery. Patients who returned the baseline questionnaire package, but did not return the follow-up questionnaire package, were reminded by regular mail or telephone two weeks after surgery. Only patients who returned both the baseline and the follow-up questionnaire packages were included into our analyses. All clinical information (e.g. ASA physical status, surgical procedure, type of anesthesia, duration of the procedure, unplanned admission and readmission) was acquired by systematic chart review.

Statistical analysis

First, missing data of potential predictor variables were imputed using multiple imputation according to the method described by Van Buuren et al.[20] To compare APSP after various types of surgery, homogenous surgical groups were created. Surgical groups were selected when they contained at least 20 procedures [11].

External validation of the prediction model

For each individual in our cohort, the predicted probability of moderate to severe APSP, defined as NRS higher than 4 [6], was computed using the regression coefficients of the previously published model [6]. To derive the regression coefficients from their tables, we computed the natural logarithm of the odds ratios that they presented. The probabilities were computed using the formula:

$P(\text{event}) = 1 / (1 + e^{(-(\text{linear predictor}))})$, in which the linear predictor is the sum of the regression coefficients multiplied by their respective predictor variable values. The intercept was not presented in the manuscript, but is necessary for the calculation of the linear predictor. Therefore, we estimated an intercept based on our cohort.

Briefly, the predicted probabilities were subsequently used to evaluate the discriminative ability of the model, and the model's calibration. The discriminative ability refers to the model's ability to discriminate between subjects who will, and those who will not develop APSP, and is expressed as the area under the receiver operating characteristic curve (AUC). The AUC ranges from 0.5 (i.e. no discriminative ability) to 1.0 (perfect discriminative ability). The calibration of the model refers to the agreement between predicted probabilities and observed frequencies of the outcome. In studies of external validation the calibration of a model is usually examined by computing the calibration in-the-large (i.e. a comparison of the average predicted probability for the whole cohort and the proportion of patients with postoperative pain) and by visually inspecting a calibration plot. Because we estimated the intercept for the model on our own data, calibration in-the-large will be spot-on. Therefore, we will confine ourselves to an inspection of the calibration plot.

Modification of the prediction model

Potential predictor variables for the modified prediction model consisted of the initial variables comprised in the previously published model before dichotomization and additional variables (i.e. ASA-level, work status, preoperative analgesic use and baseline QOL). Furthermore, type of surgery was assigned to a wide range of surgical procedures (or groups of closely related procedures). Finally, moderate to severe APSP was defined as NRS > 3 [12,13].

A multivariable logistic regression analysis was performed to estimate the regression coefficients of all variables. Only variables with a p-value < 0.1 were included in the final model. A stepwise forward

multivariable logistic regression analysis was performed to determine predictors for APSP. Only variables that were significant in more than halve of the imputed datasets were considered as significant predictors in the pooled regression model.

The development of the prediction model was based on three consecutive steps. In a first step, those variables that are easily to obtain (i.e. gender, age and surgical procedure) were included. In a second step, variables based on items which are relatively easy to obtain during the preoperative assessment (i.e. ASA-level, work status, education level, previous surgery, preoperative pain and preoperative analgesic use) were incorporated. In the third and final step, psychological variables were added to the model (i.e. expected postoperative pain by the patient, short and long term surgical fear, pain catastrophizing, optimism and preoperative QOL).

Internal validation of the modified prediction model

It is a well-known artifact that a prediction model performs considerably less well in future patients, as compared to the patients in the cohort the model was derived on. Therefore, we internally validated the model. Standard bootstrap validation was used with 1000 bootstrap samples on each of the imputed datasets [21]. Results from the bootstrap averaged over the ten datasets yielded a measure that was used to subtract from the computed AUC to obtain a conservative estimate, and a shrinkage factor used to multiply the regression coefficients by. The shrunk regression coefficients will produce less extreme results for future patients to counteract the too extreme predictions that are often observed when using a model that had not been internally validated.

All analyses were performed using SPSS version 20.0 (SPSS Inc, Chicago, IL) and R version 3.2.2.

Results

General characteristics

Patient data are given in Figure 1. A total of 1118 patients were included for the final analysis.

Baseline patient characteristics are summarized in Table 1. Patients included into the study of Gramke et al. [6] are slightly younger and less educated as compared to the patients included in our cohort. Furthermore, in our cohort, more patients had moderate to severe pain in the preoperative phase, pain medication was more often used in the week before surgery and more patients were included in the

anticipated postoperative pain level 'intermediate'. In addition, more patients received general anesthesia as compared to the patients included in the study of Gramke et al. [6].

External validation of the prediction model

The following regression models could be constructed from the results of Gramke et al. (2009) [6], after estimating intercepts specific for our cohorts:

Step 1

Predicted probalility APSP=1 / {1 + EXP (-[-2.125 + (0.956*Anticipated pain level) + (0.531*Age <45 vs 60+) - (0.357*Age 45-59 vs 60+) + (0.336*Sex)])}.

Step 2

Predicted probalility APSP=1 / {1 + EXP(-[-2.878 + (0.956*Anticipated pain level) + (0.531*Age <45 vs 60+) - (0.357*Age 45-59 vs 60+) + (0.336*Sex) + (1.131*Preoperative pain)])}.

Step 3

Predicted probalility APSP=1 / {1 + EXP(-[-3.481 + (0.956*Anticipated pain level) + (0.531*Age <45 vs 60+) - (0.357*Age 45-59 vs 60+) + (0.336*Sex) + (1.131*Preoperative pain) + (1.099*Expected pain)])}.

For each individual in our cohort, the predicted probability of APSP was computed using these formulas, leading to the AUC's shown in Table 2. The AUC's in all three steps are much lower in our validation dataset compared to the AUC's in the derivation dataset, presented by Gramke et al.[6].

For the previously published prediction model, the agreement between the predicted risk and the observed incidence of APSP applied to our data is shown in Figure 2. Although the risk-based groups lie close to the ideal 45-degree line, the relative spread of the groups around the average frequency of APSP is limited.

Modification of the prediction model

Results of the modified prediction model are shown in Table 3.. The AUC of step 1 (age, gender and surgical procedure) is 0.73. After correction for optimism (i.e. the likely performance of the model in future patients) the AUC reduced to 0.70 (Table 4).

Adding ASA status, paid work, level of education, preoperative pain and preoperative analgesic use to the regression model, the AUC improves to 0.79, with a reduction to 0.75 after correction for optimism (Table 4). Finally, the addition of expected pain, long term surgical fear and optimism, resulted in an AUC of 0.82, 0.78 after correction for optimism (Table 4).

The regression formulas for each step of the modified regression model, with a correction for optimism are the following:

Step 1

Predicted probalility APSP=1 / {1 + EXP(-[-0.85 + (Beta step 1*Surgical procedure) + (-0.02*Age) + (0.24*Sex)])}.

Step 2

Predicted probalility APSP=1 / {1 + EXP(-[-0.71 + (Beta step 2*Surgical procedure) + (-0.02*Age) + (0.18*Sex) + (-0.69*ASA 1 vs 3) + (-0.80*ASA 2 vs 3) + (0.30*Paid Work) + (0.52*Low vs High education) + (0.27*Middle vs High education) + (0.84*Preoperative pain) + (0.44*Preoperative analgesic use)])}.

Step 3

Predicted probalility APSP=1 / {1 + EXP(-[-0.30 + (Beta step 3*Surgical procedure) + (-0.17*Age) + (0.15*Sex) + (-0.73*ASA 1 vs 3) + (-0.89*ASA 2 vs 3) + (0.33*Paid Work) + (0.49*Low vs High education) + (0.28*Middle vs High education) + (0.60*Preoperative pain) + (0.37*Preoperative analgesic use) + (0.67*Expected pain) + (0.03*Long term surgical fear) + (-0.07*Optimism)])}.

The regression coefficients or beta's for surgical procedure for the three steps of the regression formula are given in table 4.

For example, if a 60-year old male patient received anal surgery, using the regression formula of step 1, his individual probability of APSP will be:

Predicted probalility APSP=1 / {1 + EXP(-[-0.85 + **(1.66*1)** + (-0.02*60) + (0.24*1)])} = 0.46 = 46%.'

If, in addition, we know that his ASA status is 2, he has a paid job, a low level of education, experienced preoperative pain and used analgesics prior to the surgical procedure, his individual probability of APSP, using the regression formula of step 2, will be:

Predicted probability APSP=1 / {1 + EXP(- [-0.71 + (1.41*1) + (-0.02*60) + (0.18*1) + (-0.80*1) + (0.30*1) + (0.52*1) + (0.84*1) + (0.44*1)])} = 0.73 = 73%.

Calibration curves for each consecutive step of the prediction model (i.e. step 1 through 3) are shown in Figure 3. These curves indicate good calibration of the prediction model because the risk-based groups are all close to the ideal 45-degree line and they cover the whole range of probabilities (i.e. between 0 and 1), especially for step 2 and 3.

Discussion

In the present study, we externally validated and subsequently modified a previously developed model [6] to preoperatively predict the risk of moderate to severe APSP in surgical outpatients on the fourth postoperative day. Finally, we internally validated the modified prediction model.

The predictive accuracy of the 3-step model described by Gramke et al. [6] was substantially lower in our validation dataset of 1118 patients than in the original dataset that was used to develop this model. The ability of this model to discriminate between the presence and absence of APSP was poorer with, after the 3rd step, an AUC of 0.81 in the derivation dataset [6] and an AUC of only 0.72 in the validation dataset (Table 2). Furthermore, the calibration plot of the 3rd step of this model (Figure 2) shows a risk prediction that is too extreme, i.e. an underestimation of the predicted low risks and a distinct overestimation of the predicted high risks.

Modification of the original model consisted of assigning type of surgery to a wide range of surgical procedures (or groups of closely related procedures), defining moderate to severe pain as NRS higher than 3, use of continuous variables instead of dichotomized variables and testing the predictive power of additional variables (i.e. ASA-level, work status, preoperative analgesic use and baseline QOL). Our modified model showed that preoperative pain, patient derived expected pain and certain types of surgery are the best predictors of moderate to severe APSP on the fourth day after day surgery (Table 3). Other predictors are younger age, higher ASA status, paid work, low level of education, preoperative analgesic use, long term surgical fear and pessimism. Moreover, our modified 3-step model is able to discriminate between patients with and without moderate to severe APSP with an

AUC of 0.82 and after correction for optimism still an AUC of 0.78 (Tables 3 and 4). The calibration curves shown in Figure 3 indicate good calibration of the modified model.

When applied to new individuals, the performance of a prediction model is generally lower than the performance observed in the population from which the model was initially developed [22]. Poor performance in new patients may be due to overfitting of the model and can also arise from differences in patient characteristics, distribution of predictor values between both datasets, methods of measurement and healthcare system [7]. In our study, the flattened slope of the calibration plot of the original model (figure 2) and the observed decrease in AUC (table 2) are clear signals of overfitting the model and optimism in the performance parameters [23]. Furthermore, pain intensity was measured using the VAS in the derivation study in contrast to the NRS in the validation study. Finally, our patient cohort included more patients with preoperative pain and an 'intermediate' level of anticipated postoperative pain and they were slightly older and more highly educated (Table 1) as compared to the cohort used by Gramke et al [7]. These differences can be explained by the recent evolution of day surgery towards more complex surgical procedures on older and higher risk patients [24].

Various predictors of postoperative pain have been reported in literature. Based on a systematic review preoperative pain, anxiety, psychological distress (i.e. pain catastrophizing, pessimism, depression), younger age and type of surgery were reported to be the five most significant predictive factors for postoperative pain [25]. This systematic review did not include preoperative expectations of postoperative pain (by the patient) as a possible predictor. Nevertheless, a positive correlation between preoperative expectations of pain and the occurrence of postoperative pain has been reported in literature [6,26,27]. Preoperative expectations by the patient on postoperative pain are influenced by many factors including previous experiences with surgery or other traumatic injuries, the individual memory and psychological profile of the patient [6]. Unlike other studies, pain catastrophizing was not found to be a significant predictor in our model. The predictive value of pain catastrophizing may have been reduced in our model by the inclusion of another psychological predictor 'preoperative expectations of postoperative pain' since these two variables seem to be associated [28]. Our model also showed higher ASA status to be a predictor of APSP and a similar correlation was reported by Caumo et al. [29].

Furthermore, our analysis showed that patients with a paid job reported a higher APSP as compared to patients without. It can be hypothesized that patients with a paid job desire longer sick leave and therefore tend to overestimate their pain levels.

Limited data exist on the effect of educational level on APSP [6,30]. In the present study, we report that a low level of education is a significant predictor for APSP and a similar correlation was found in two previous studies [6,30]. This association might be related to differences in the ability to cope with pain [6].

The observed correlation between preoperative analgesic use and APSP is in line with a previous study [31]. Although the relation between preoperative analgesic use and APSP is not clear, three possible mechanisms might be involved: tachyphylaxis, opioid-induced hyperalgesia (in case of preoperative use of opioids) and neuroplastic changes in the spinal cord due to chronic noxious input [27].

The modified prediction model can be valuable when implemented in the regular preoperative anesthesia evaluation of the outpatient. Identification of patients at high risk for moderate to severe APSP enables physicians to plan a tailor-made effective postoperative analgesic regimen and a more comprehensive follow-up program for these patients. In practice, this includes use of multi-modal analgesic techniques, regular telephone follow-up and even planned overnight stay. Moreover, it enables better patient information provision and adequate use of resources for selected patients with increased risk profile. Implementation of the modified model in daily practice can be achieved with the development of a convenient medical software application. After input of patient demographics, other patient-related predictive factors and data on type of surgery, this application can easily calculate the risk for APSP with the regression formula of the modified prediction model.

This study also has some limitations. Firstly, Gramke et al. [6] assessed postoperative pain at the day of operation and at postoperative days 1 to 4. In contrast, our validation study only assessed postsurgical pain at the fourth postoperative day. Hence, validation of the predictive model of Gramke et al. in the present study is limited to the fourth postsurgical day. Still, the discriminative power of the predictive model of Gramke et al. did not vary over the four postoperative days [6]. Secondly, pain intensity was measured using the VAS in the derivation study in contrast to the NRS in the validation study. However, it has been proven that VAS and NRS scores correspond well [32,33]. Furthermore, the

generalizability of a prediction model can only be proven if this model has been tested in a more diverse setting [34]. Thirdly, we performed a temporal validation (dataset for validation was collected at the same centre). As a result, the extrapolation of the predictive performance of the model to other institutes or countries cannot be proven [22]. Still, the variation between the two different datasets has been enlarged by the fact that the validation has been performed by a different team with overlapping authors and by the large time frame between data collection (more than 6 years). More specifically, within this time frame, different strategies in postsurgical pain therapy have been implemented and improvements in surgical techniques and antiemetic therapy made it possible to perform more complex surgical procedures in an older and higher risk patient population in the outpatient setting[24]. Our results confirm these recent developments. Finally, the complexity of this modified model can impede his implementation in clinical practice. However, as mentioned above, a convenient medical software application can overcome this obstacle.

In conclusion, we could not validate the use of the prediction model of Gramke et al. on a large cohort of outpatients since both discrimination and calibration were considerably less than expected. Internal validation of our modified version of this model however shows promising results with good discrimination and calibration. In this modified model, preoperative pain, patient derived expected pain and certain types of surgery are the best predictors of moderate to severe APSP after day surgery. Other predictors are younger age, higher ASA status, paid work, low level of education, preoperative analgesic use, long term surgical fear and pessimism.

References

1. Wu CL, Berenholtz SM, Pronovost PJ, et al. Systematic review and analysis of postdischarge symptoms after outpatient surgery. Anesthesiology. Apr 2002;96(4):994-1003.
2. Rawal N. Postoperative pain treatment for ambulatory surgery. Best practice & research. Clinical anaesthesiology. Mar 2007;21(1):129-148.
3. Gramke HF, de Rijke JM, van Kleef M, et al. The prevalence of postoperative pain in a cross-sectional group of patients after day-case surgery in a university hospital. The Clinical journal of pain. Jul-Aug 2007;23(6):543-548.
4. McGrath B, Elgendy H, Chung F, et al. Thirty percent of patients have moderate to severe pain 24 hr after ambulatory surgery: a survey of 5,703 patients. Canadian journal of anaesthesia = Journal canadien d'anesthesie. Nov 2004;51(9):886-891.
5. Beauregard L, Pomp A, Choiniere M. Severity and impact of pain after day-surgery. Canadian journal of anaesthesia = Journal canadien d'anesthesie. Apr 1998;45(4):304-311.
6. Gramke HF, de Rijke JM, van Kleef M, et al. Predictive factors of postoperative pain after day-case surgery. The Clinical journal of pain. Jul-Aug 2009;25(6):455-460.

7. Altman DG, Vergouwe Y, Royston P, et al. Prognosis and prognostic research: validating a prognostic model. BMJ. 2009;338:b605.
8. Siontis GC, Tzoulaki I, Castaldi PJ, et al. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. Journal of clinical epidemiology. Jan 2015;68(1):25-34.
9. Collins GS, de Groot JA, Dutton S, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. BMC medical research methodology. 2014;14:40.
10. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. Jan 2010;21(1):128-138.
11. Gerbershagen HJ, Aduckathil S, van Wijck AJ, et al. Pain intensity on the first day after surgery: a prospective cohort study comparing 179 surgical procedures. Anesthesiology. Apr 2013;118(4):934-944.
12. Dihle A, Helseth S, Paul SM, et al. The exploration of the establishment of cutpoints to categorize the severity of acute postoperative pain. The Clinical journal of pain. Sep 2006;22(7):617-624.
13. Gerbershagen HJ, Rothaug J, Kalkman CJ, et al. Determination of moderate-to-severe postoperative pain on the numeric rating scale: a cut-off point analysis applying four different methods. Br J of Anaesth. Oct 2011;107(4):619-626.
14. Van Agt HM, Essink-Bot ML, Krabbe PF, et al. Test-retest reliability of health state valuations collected with the EuroQol questionnaire. Social science & medicine. Dec 1994;39(11):1537-1544.
15. Peters ML, Sommer M, de Rijke JM, et al. Somatic and psychologic predictors of long-term unfavorable outcome after surgical intervention. Annals of surgery. Mar 2007;245(3):487-494.
16. Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: Development and validation. Psychological Assessment. 1995;7:524-532.
17. Scheier MF, Carver CS, Bridges MW. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. Journal of personality and social psychology. Dec 1994;67(6):1063-1078.
18. Theunissen M, Peters ML, Schouten EG, et al. Validation of the surgical fear questionnaire in adult patients waiting for elective surgery. PloS one. 2014;9(6):e100225.
19. Stessel B, Fiddelers AA, Joosten EA, et al. Prevalence and Predictors of Quality of Recovery at Home After Day Surgery. Medicine. Sep 2015;94(39):e1553.
20. Van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Statistics in medicine. Mar 30 1999;18(6):681-694.
21. Steyerberg EW. Clinical Prediction Models. A Practical Approach to Development, Validation, and Updating. . Springer Science+Business Media, LCC. 2009.
22. Moons KG, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart (British Cardiac Society). May 2012;98(9):691-698.
23. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in medicine. Feb 28 1996;15(4):361-387.
24. White PF, White LM, Monk T, et al. Perioperative care for the older outpatient undergoing ambulatory surgery. Anesthesia and analgesia. Jun 2012;114(6):1190-1215.
25. Ip HY, Abrishami A, Peng PW, et al. Predictors of postoperative pain and analgesic consumption: a qualitative systematic review. Anesthesiology. Sep 2009;111(3):657-677.
26. Mamie C, Bernstein M, Morabia A, et al. Are there reliable predictors of postoperative pain? Acta anaesthesiologica Scandinavica. Feb 2004;48(2):234-242.
27. Sommer M, de Rijke JM, van Kleef M, et al. Predictors of acute postoperative pain after elective surgery. The Clinical journal of pain. Feb 2010;26(2):87-94.
28. Sommer M, Geurts JW, Stessel B, et al. Prevalence and predictors of postoperative pain after ear, nose, and throat surgery. Archives of otolaryngology--head & neck surgery. Feb 2009;135(2):124-130.

29. Caumo W, Schmidt AP, Schneider CN, et al. Preoperative predictors of moderate to intense acute postoperative pain in patients undergoing abdominal surgery. Acta Anaesthesiol Scand. Nov 2002;46(10):1265-1271.
30. Taenzer P, Melzack R, Jeans ME. Influence of psychological factors on postoperative pain, mood and analgesic requirements. Pain. Mar 1986;24(3):331-342.
31. Aubrun F, Valade N, Coriat P, et al. Predictive factors of severe postoperative pain in the postanesthesia care unit. Anesthesia and analgesia. May 2008;106(5):1535-1541, table of contents.
32. Hjermstad MJ, Fayers PM, Haugen DF, et al. Studies comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for assessment of pain intensity in adults: a systematic literature review. Journal of pain and symptom management. Jun 2011;41(6):1073-1093.
33. Ferreira-Valente MA, Pais-Ribeiro JL, Jensen MP. Validity of four pain intensity rating scales. Pain. Oct 2011;152(10):2399-2404.
34. van den Bosch JE, Kalkman CJ, Vergouwe Y, et al. Assessing the applicability of scoring systems for predicting postoperative nausea and vomiting. Anaesthesia. Apr 2005;60(4):323-331.

Figure 1: Flowchart depicting the inclusion and exclusion.

Figure 2. Calibration curves of the external validation of the previously published prediction model (Gramke et al. 2009). Triangles indicate the observed frequency of moderate to severe APSP per decile of predicted risk. The solid line shows the relation between observed outcomes and predicted risks. Ideally, this line equals the dotted line that represents perfect calibration. The histogram on the x-axis shows the distribution of predicted risks in the external validation data.

Figure 3. Calibration curves of the modified prediction model. Triangles indicate the observed frequency of moderate to severe APSP per decile of predicted risk. The solid line shows the relation between observed outcomes and predicted risks. Ideally, this line equals the dotted line that represents perfect calibration. The histogram on the x-axis shows the distribution of predicted risks.

Table 1: Baseline characteristics of the validation dataset (Stessel et al.) and of the derivation dataset (Gramke et al.). Data are presented as absolute number (N) and percentage, or as mean and standard deviation (SD).

	N (%) / Mean (SD) Stessel et al.	**N (%) / Mean (SD) Gramke et al. (2009)**
Age	52.5 (14.3)	49.3 (16.9)
< 45 years	340 (30)	240 (37)
45 – 59 years	387 (35)	232 (36)
> 59 years	391 (35)	176 (27)
Gender		
Male	505 (45)	281 (43)
Female	613 (55)	367 (57)
Educational background		
Elementary school	356 (32)	221 (34)
Intermediate	430 (38)	247 (38)
Higher degree, university	319 (29)	170 (26)
Information missing	13 (1)	10 (2)
Preoperative pain		
VAS > 10mm/NRS > 1	592 (53)	138 (21)
VAS > 30mm/NRS > 3	424 (38)	71 (11)
Analgesic use before operation		
Acetaminophen	132 (12)	39 (6)
NSAIDs	94 (8)	43 (7)
Weak opioids	41 (4)	12 (2)
None	831 (74)	554 (85)
Anticipated postoperative pain level, Based on the type of surgery		
Minor	647 (580)	452 (70)
Intermediate	469 (42)	196 (30)
Type of anesthesia		
General	933 (84)	400 (62)
Regional	182 (16)	248 (38)

VAS = Visual Analog Scale, NRS = Numeric Rating Scale, NSAID = Non-steroidal anti-inflammatory drug

Table 2. Discriminative ability of the previously published model (Gramke et al.) [7] in the validation dataset (Stessel et al.) versus the discriminative ability of this model in the original dataset (Gramke et al.) [7].

Regression model	Stessel et al. AUC	Gramke et al. AUC
Step 1	0.58	0.72
Step 2	0.66	0.78
Step 3	0.72	0.81

AUC = Area under the Receiver Operating Characteristic curve.

Table 3. Results of the modified prediction model: associations between predictor variables and acute postsurgical pain

Independent variable	n	Step 1 OR (95% CI); AUC = 0.73	Step 2 OR (95% CI); AUC = 0.79	Step 3 OR (95% CI); AUC = 0.82
Step 1				
Age		0.98 (0.97 – 0.99)	0.98 (0.97 – 0.99)	0.98 (0.97 – 0.99)
Gender (male vs female)		1.36 (0.99 – 1.87)	1.25 (0.88 – 1.77)	1.20 (0.83 – 1.74)
Surgical procedure	49			
1. Diagnostic laryngoscopy		Reference	Reference	Reference
2. Anal surgery	51	8.62 (3.09 – 24.07)	5.73 (2.01 – 16.39)	5.02 (1.68 – 14.97)
3. Cataract / vitrectomy	61	0.69 (0.18 – 2.64)	0.69 (0.18 – 2.65)	0.72 (0.18 – 2.88)
4. Dupuytren fasciotomy	32	7.26 (2.34 – 22.50)	4.83 (1.50 – 15.49)	3.74 (1.11 – 12.61)
5. Umbilical / epigastric / cicatricalic hernia repair	26	1.09 (0.25 – 4.82)	0.69 (0.15 – 3.12)	0.81 (0.17 – 3.84)
6. Hysteroscopy	47	0.98 (0.29 – 3.32)	0.78 (0.22 – 2.74)	0.81 (0.23 – 2.87)
7. Laparoscopic sterilisation / ovariectomy	30	1.46 (0.42 – 5.10)	1.30 (0.36 – 4.74)	1.04 (0.27 – 3.98)
8. Lumpectomy	42	0.39 (0.07 – 2.10)	0.35 (0.07 – 1.88)	0.29 (0.05 – 1.62)
9. (Sub)cutaneous surgery	76	2.06 (0.74 – 5.70)	1.30 (0.45 – 3.73)	1.25 (0.42 – 3.71)
10. Strabismus surgery	20	1.83 (0.45 – 7.43)	1.81 (0.42 – 7.75)	1.74 (0.37 – 8.16)
11. Tendon / bursa / fascia surgery	57	6.16 (2.24 – 16.94)	2.69 (0.95 – 7.65)	2.40 (0.81 – 7.16)
12. Scrotal surgery	20	4.93 (1.37 – 17.71)	3.70 (0.99 – 13.82)	4.09 (1.03 – 16.20)
13. Nose – sinus / polyp / septum surgery	29	2.54 (0.77 – 8.35)	1.63 (0.47 – 5.66)	1.10 (0.29 – 4.24)
14. Tympanoplasty /	31	1.34 (0.37 – 4.89)	1.23 (0.33 –	1.20 (0.31 –

stapedectomy / ossicular chain reconstruction			4.60)	4.64)
15. Brachytherapy	32	0.79 (0.15 – 4.29)	0.73 (0.13 – 4.03)	0.56 (0.10 – 3.17)
16. Dental surgery	24	9.76 (2.91 – 32.76)	7.03 (1.97 – 25.17)	5.33 (1.42 – 20.01)
17. Arthroscopy knee / meniscectomy	146	4.61 (1.82 – 11.64)	1.73 (0.66 – 4.54)	1.59 (0.58 – 4.31)
18. Mamma reconstruction / implants	21	1.08 (0.24 – 4.84)	0.84 (0.18 – 3.95)	0.82 (0.17 – 3.86)
19. Mamma reduction / mastectomy	24	2.06 (0.58 – 7.36)	1.62 (0.42 – 6.16)	1.21 (0.31 – 4.72)
20. Hardware removal	48	5.33 (1.89 – 15.06)	2.80 (0.95 – 8.25)	2.89 (0.93 – 8.98)
21. Inguinal hernia repair	72	7.19 (2.61 – 19.79)	4.46 (1.58 – 12.57)	4.46 (1.51 – 13.16)
22. Laparoscopic cholecystectomy	41	1.50 (0.46 – 4.94)	0.69 (0.20 – 2.39)	0.63 (0.18 – 2.26)
23. Shoulder surgery	41	9.95 (3.43 – 28.82)	3.98 (1.32 – 12.03)	3.60 (1.16 – 11.24)
24. Bone surgery	57	5.02 (1.82 – 13.84)	2.59 (0.90 – 7.45)	2.25 (0.75 – 6.74)
25. Mastoidectomy / CAT / BAHA	41	1.55 (0.45 – 5.28)	1.09 (0.31 – 3.84)	0.85 (0.22 – 3.32)
Step 2				
ASA status				
1 vs. 3			0.42 (0.20 – 0.91)	0.41 (0.18 – 0.92)
2 vs. 3			0.37 (0.18 – 0.78)	0.34 (0.15 – 0.74)
Paid work				
Yes vs. No			1.45 (1.02 – 2.04)	1.51 (1.05 – 2.17)
Level of education				
Low vs. High			1.90 (1.24 – 2.91)	1.81 (1.16 – 2.85)
Middle vs. High			1.39 (0.94 – 2.05)	1.42 (0.94 – 2.14)
Preoperative pain				
Yes vs. No			2.82 (1.96 – 4.07)	2.09 (1.41 – 3.08)
Preoperative analgesic use				
Yes vs. No			1.73 (1.20 – 2.49)	1.57 (1.07 – 2.29)
Step 3				
Expected pain				2.26 (1.58 – 3.23)
Surgical fear – long term (high vs low)				1.04 (1.02 – 1.06)
Optimism				0.93 (0.87 – 0.99)

OR = odds ratio, AUC = Area under the Receiver Operating Characteristic curve

Table 4: Regression coefficients (= beta´s) of the modified prediction model corrected for overfitting (i.e. they were penalized, or shrunk towards 0, by multiplying them with the shrinkage factor resulting from the bootstrap validation). The coefficients can be used to compute an individuals' probability of acute postsurgical pain.

Independent variable	**Step 1 Coefficients after shrinkage***	**Step 2 Coefficients after shrinkage***	**Step 3 Coefficients after shrinkage***
Step 1			
Constant	-0.85	-0.71	-0.30
Age	-0.02	-0.02	-0.17
Gender (male vs female)	0.24	0.18	0.15
Surgical procedure (vs. Diagnostic laryngoscopy)			
1. Anal surgery	1.66	1.41	1.31
2. Cataract / vitrectomy	-0.29	-0.31	-0.27
3. Dupuytren fasciotomy	1.53	1.26	1.07
4. Umbilical / epigastric / cicatricalic hernia repair	0.06	-0.30	-0.17
5. Hysteroscopy	-0.02	-0.20	-0.17
6. Laparoscopic sterilisation / ovariectomy	0.29	0.22	0.03
7. Lumpectomy	-0.72	-0.85	-1.01
8. (Sub)cutaneous surgery	0.56	0.21	0.18
9. Strabismus surgery	0.46	0.48	0.45
10. Tendon / bursa / fascia surgery	1.41	0.80	0.72
11. Scrotal surgery	1.23	1.05	1.15
12. Nose – sinus / polyp / septum surgery	0.72	0.39	0.08
13. Tympanoplasty / stapedectomy / ossicular chain reconstruction	0.22	0.17	0.15
14. Brachytherapy	-0.18	-0.26	-0.47
15. Dental surgery	1.76	1.57	1.36
16. Arthroscopy knee / meniscectomy	1.18	0.44	0.37
17. Mamma reconstruction / implants	0.05	-0.14	-0.16
18. Mamma reduction / mastectomy	0.56	0.39	0.15
19. Hardware removal	1.29	0.83	0.86
20. Inguinal hernia repair	1.52	1.20	1.22
21. Laparoscopic	0.32	-0.30	-0.37

cholecystectomy			
22. Shoulder surgery	1.78	1.11	1.04
23. Bone surgery	1.24	0.76	0.66
24. Mastoidectomy / CAT / BAHA	0.34	0.07	-0.13
Step 2			
ASA status			
1 vs. 3		-0.69	-0.73
2 vs. 3		-0.80	-0.89
Paid work			
Yes vs. No		0.30	0.33
Level of education			
Low vs. High		0.52	0.49
Middle vs. High		0.27	0.28
Preoperative pain			
Yes vs. No		0.84	0.60
Preoperative analgesic use			
Yes vs. No		0.44	0.37
Step 3			
Expected pain			0.67
Surgical fear – long term (high vs low)			0.03
Optimism			-0.07

*Beta's are corrected for overfitting with the following Shrinkage factors (SF) derived from the bootstrap internal validation: SF Model 1 = 0.7725, SF Model 2 = 0.8052, SF Model 3 = 0.8127. The optimism-corrected area's under the Receiver operating Characteristic curves (AUC) are, respectively, 0.70, 0.75, and 0.78, for Models 1, 2, and 3.

Figure 1.

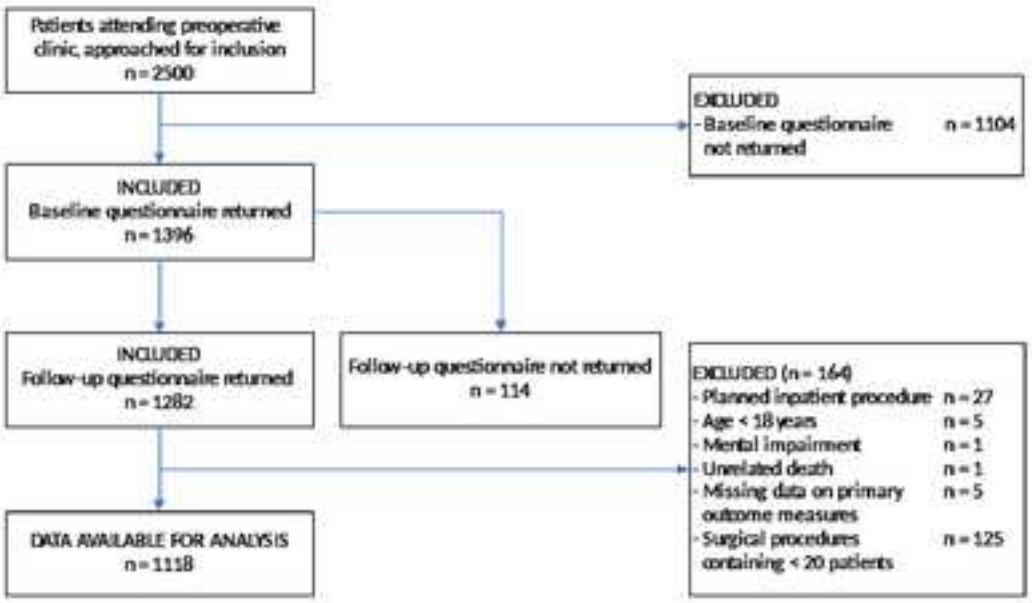
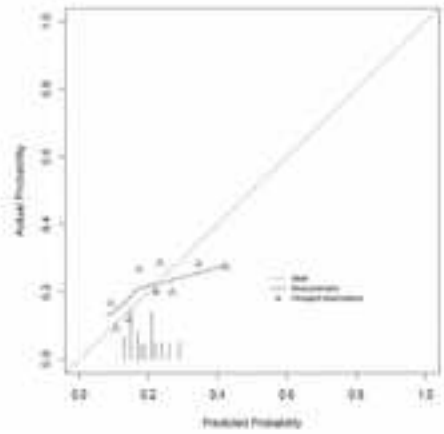
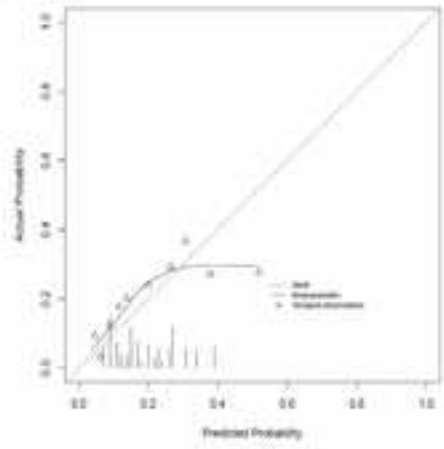


Figure 2.

Step 1



Step 2



Step 3

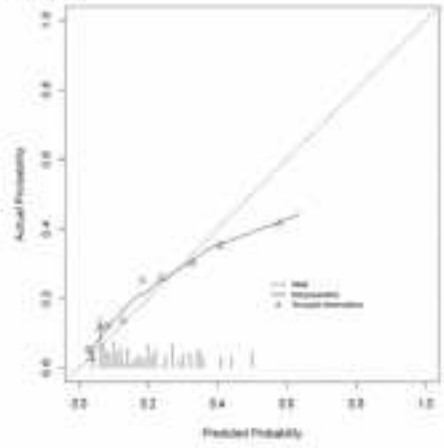
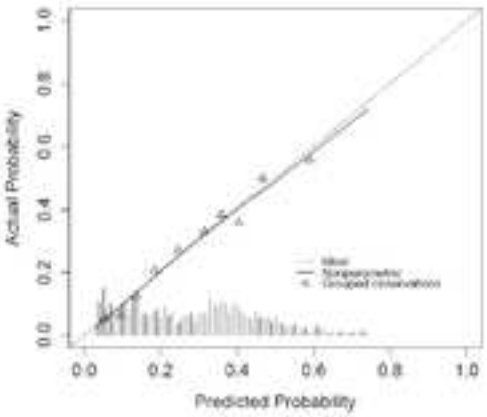
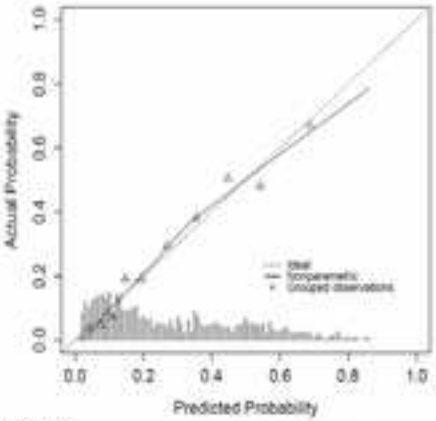


Figure 3.

Step 1



Step 2



Step 3